available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel

Incorporating anthropogenic variables into a species distribution model to map gypsy moth risk

Christopher D. Lippitt[a,b,*], *John Rogan*[a], *James Toledano*[b], *Florencia Sangermano*[a,b], *J. Ronald Eastman*[a,b], *Victor Mastro*[c], *Alan Sawyer*[c]

[a] *Graduate School of Geography, Clark University, 950 Main St., Worcester, MA 01610, USA*
[b] *Clark Labs, 921 Main St., Worcester, MA 01610, USA*
[c] *United States Department of Agriculture, Animal and Plant Health Inspection Service, PPQ-PSDEL, Bldg. 1398, W. Truck Rd., Otis Air National Guard Base, MA 02542, USA*

ARTICLE INFO

Article history:
Received 21 May 2007
Received in revised form
30 July 2007
Accepted 7 August 2007
Published on line 17 September 2007

Keywords:
Species distribution modeling
Anthropogenic
Neural network
Risk
Invasive species

ABSTRACT

This paper presents a novel methodology for multi-scale and multi-type spatial data integration in support of insect pest risk/vulnerability assessment in the contiguous United States. Probability of gypsy moth (*Lymantria dispar* L.) establishment is used as a case study. A neural network facilitates the integration of variables representing dynamic anthropogenic interaction and ecological characteristics. Neural network model (back-propagation network [BPN]) results are compared to logistic regression and multi-criteria evaluation via weighted linear combination, using the receiver operating characteristic area under the curve (AUC) and a simple threshold assessment. The BPN provided the most accurate infestation-forecast predictions producing an AUC of 0.93, followed by multi-criteria evaluation (AUC = 0.92) and logistic regression (AUC = 0.86) when independently validating using post model infestation data. Results suggest that BPN can provide valuable insight into factors contributing to introduction for invasive species whose propagation and establishment requirements are not fully understood. The integration of anthropogenic and ecological variables allowed production of an accurate risk model and provided insight into the impact of human activities.
© 2007 Elsevier B.V. All rights reserved.

1. Introduction

Species distribution models (SDMs) are playing an ever-increasing role in understanding the current and potential future distribution of flora and fauna. SDMs relate plant and animal distribution to ecological variables that contribute to their persistence and/or propagation (Guisan and Zimmermann, 2000). We present a novel methodology for integrating ecological and anthropogenic data in distribution models to support insect pest risk assessment in the contiguous United States (US). The gypsy moth (*Lymantria dispar* L.), an invasive species in the US, is used as a case study to compare the performance of expert, parametric, and neural network models for integrative risk assessment.

There are approximately 50,000 invasive species in the United States (Pimentel et al., 1999) collectively affecting every state and territory (Bergman et al., 2000). Pimentel et al. (1999) estimate total invasive species damage to be approximately $138 billion per annum; $2.1 billion of which is attributed to forest pests such as the gypsy moth. The gypsy moth alone has defoliated millions of hectares of valuable timber species (Gerardi and Grimm, 1979) causing millions of dollars of dam-

* *Corresponding author at*: San Diego State University, Department of Geography, 5500 Campanile Dr., San Diego, CA 92182-4493, USA. Tel.: +1 508 849 2322.
E-mail address: Lippitt@rohan.sdsu.edu (C.D. Lippitt).
0304-3800/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2007.08.005

age each year (Leuschner et al., 1996) with a host of ecological problems (Gottschalk, 1993). Every year, more than 250,000 ha of US forest are treated in an attempt to minimize gypsy moth defoliation impacts (USDA Forest Service, 1992) and there is concern that it may be spreading to areas previously believed to be uninhabitable (Allen et al., 1993). If uncontrolled, it is likely the gypsy moth will extend its range to most of the contiguous US and southern Canada (Liebhold et al., 1992; Sharov et al., 1997).

The United States Department of Agriculture's Animal and Plant Health Inspection Service (APHIS), the agency charged with the detection and mitigation of gypsy moth, requires an improved decision support tool to aid the prediction of gypsy moth introduction, establishment, and spread for the contiguous United States. Current gypsy moth decision support consists of non-spatial, unsystematic, estimations by regional managers (USDA, 2001). Geographic Information Science (GIScience) and technology offer the capability to characterize insect infestation probability in a spatially explicit, accurate, and replicable method; a function vital to managers charged with the efficient distribution of limited detection and mitigation resources over large spatial extents (Byers et al., 2002; Stohlgren and Schnase, 2006).

Modeling gypsy moth risk with commonly used techniques, however, presents two challenges: ecological variables typically included in SDMs do not account for anthropogenic impacts on the response variable; and methods traditionally used to model spatial variables require *a priori* definition of variable relationships and/or violate basic statistical assumptions of independence and/or linearity (Gahegan, 2003). Machine learning (e.g., neural network) methods allow the characterization of models containing non-linear relationships among, and between predictor variables without the explicit definition of those relationships (Foody, 1995; Lek et al., 1996; Lek and Guegan, 1999).

This research predicts gypsy moth infestation risk in non-infested counties of the contiguous US to: (1) assess the capability of an automated artificial neural network (ANN) to integrate environmental and anthropogenic variables for predictive modeling in comparison to other commonly employed SDM techniques; and (2) improve upon previously developed gypsy moth infestation risk schemes through the incorporation of anthropogenic variables.

2. Background

2.1. Gypsy moth ecology

Since its introduction in Massachusetts (i.e., 1868 or 1869) the gypsy moth has expanded its range to include the entire northeastern portion of the US including portions of Virginia, West Virginia, Ohio, Indiana, North Carolina and Michigan (Liebhold et al., 1989, 1996). Gypsy moth still only occupies 23% of the estimated 607 million ha in its potential range (US only) (Liebhold et al., 1997a; Morin et al., 2005). One of the primary reasons for the gypsy moth's successful propagation is that it is known to utilize nearly 300 tree species as primary hosts (Leonard, 1981; Liebhold et al., 1995). Its ability to establish and persist, however, varies among different tree species (Herrick and Gasner, 1986). Table 1 provides a summary of predominant gypsy moth host species. The gypsy moth's preferred host species include many of the most prevalent deciduous tree species in the US. Several of the states containing the highest amount of highly susceptible forest are not currently infested (Liebhold et al., 1997b).

Table 1 – Most common gypsy moth hosts (listed in descending abundance) in the contiguous United States (adapted from Liebhold et al., 1997a,b)

Common name	Scientific name	Total basal area 100 million ft/acre
White oak	*Quercus alba*	14.3
Sweetgum	*Liquidambar styraciflua*	11.6
Quaking aspen	*Populus tremuloides*	10.1
Northern red oak	*Quercus rubra*	9.62
Black oak	*Quercus velutina*	7.31
Chestnut oak	*Quercus prinus*	6.84
Post oak	*Quercus stellata*	5.47
Water oak	*Quercus nigra*	4.34
Paper birch	*Betula papyrifera*	3.81
Southern red oak	*Quercus falcata*	3.75
Scarlet oak	*Quercus coccinea*	3.31
American basswood	*Tilia americana*	2.41
Western larch	*Larix occidentalis*	2.40
Laurel oak	*Quercus laurifolia*	1.94
Bigtooth aspen	*Populus grandidentata*	1.90
Tanoak	*Lithocarpus densiflorus*	1.64
Willow oak	*Quercus phellos*	1.49
California red oak	*Quercus kelloggii*	1.45
Eastern hophornbeam	*Ostrya virginiana*	1.26
Canyon live oak	*Quercus chrysolepis*	1.14

Female *Lymantria dispar* (L), the species of gypsy moth found in the U.S., are not flight capable, thus limiting their natural migration to <1–2 km per annum. However, potential egg mass substrate vectors include vehicles, campers, trailers, boats, lawn furniture, swing sets, barbecue grills, tarps, etc. (USDA, 2001). When people transport substrate materials, either during household moves or vacations, they may carry gypsy moths either in the pupal or egg stage. Therefore, movement of people, vehicles, and household goods from infested areas to non-infested areas is the principal mechanism for long-range dispersal of the gypsy moth (USDA, 2001). The characterization of gypsy moth introduction probability and subsequent calculation of infestation risk, therefore requires the incorporation of human (i.e., probable gypsy moth) movement data.

2.2. Species distribution modeling

SDMs relate species distribution observations to environmental predictor variables (i.e., gradients) based on statistically or theoretically-derived response functions (Guisan and Zimmermann, 2000). Austin (1980, 2002) defined three types of environmental gradients (i.e., variables) for the prediction of species distribution; *resource*, *direct*, and *indirect* gradients. Resource gradients address matter and energy consumed by plants or animals (e.g., nutrients, water, light for plants, food). Direct gradients are environmental parameters that

have physiological importance, but are not consumed (e.g., temperature, pH). Indirect gradients are variables that have no direct physiological relevance for a species' persistence and/or propagation (e.g., slope, aspect, elevation, topographic position, habitat type, geology), and are often descriptive of several direct and/or resource gradients. Anthropogenic activities influence plant and animal species in ways that have physiological effects (e.g., harvest, transportation) and in ways that do not (e.g., disturbance, habitat preservation). Anthropogenic variables, therefore, can be described as direct or indirect gradients, under the Austin (1980, 2002) classification.

SDMs are often developed for the estimation of environmental risk (e.g., Araujo and Williams, 2000; Araujo et al., 2002; Ferrier, 2002; Vander Zanden et al., 2004), a variable frequently influenced or directly caused by anthropogenic activities (e.g., climate change, development, resource extraction, invasive species, competition). The primary goal of coarse scale (i.e., national-continental) modeling of gypsy moth infestation risk, to date, has been to gain a better understanding of host species abundance in order to estimate the total area of potential infestation in the presence of an introduction. For example, Leibhold et al. (1997b, p. 20) define susceptibility (i.e., risk) as "the probability or frequency of defoliation given an established gypsy moth population". Accordingly, several studies have mapped host abundance in an attempt to estimate the total area at risk to infestation (e.g., Liebhold et al., 1997b; Morin et al., 2005). Mapping host abundance as a proxy for risk, however, ignores current ecological theory on gypsy moth movement and subsequently overestimates the area at risk to gypsy moth infestation, which could result in overspending with respect to detection and mitigation strategies (Morin et al., 2005). The calculation of gypsy moth infestation risk thus requires the inclusion of anthropogenic information on introduction probability.

A limited number of studies have incorporated anthropogenic variables into SDMs (e.g., Austin et al., 1996 [building density, road length], Osborne et al., 2001 [disturbance], Cumming, 2002 [political regions], Suárez-Seoane et al., 2002 [roads, towns]). Anthropogenic variables influence the distribution of species (Austin et al., 1996; Osborne et al., 2001) and therefore must be considered for inclusion in SDMs. Species-anthropogenic variable relationships, however, are likely to be non-linear and to exhibit strong interaction with some ecological variables. The incorporation of interaction and non-linear variables into SDMs will require the use of nonparametric modeling techniques (Ozesmi et al., 2006). The inclusion of anthropogenic variables and application of modern nonparametric statistical techniques represent rudimentary steps toward the development of statistically rigorous models rooted in sound ecological theory, which remains the fundamental benchmark for the discipline of species distribution modeling (Austin, 2002; Guisan and Thuiller, 2005; Guisan et al., 2006).

2.3. *Modeling techniques*

Decision support (i.e., resource allocation optimization) modeling has been predominantly limited to deductive techniques based on *expert* opinion (expert systems, e.g., multi-criteria evaluation) requiring *a priori* understanding of predictor/response variable relationships (Eastman et al., 1993). Inductive (i.e. empirical) techniques (e.g., logistic regression), however, offer the ability to model phenomena for which predictor–response variable relationships are not fully understood (Guisan and Zimmermann, 2000). Non-linear relationships and inherent spatial dependence within, among, and between predictor variables, however, violate assumptions of conventional statistical theory and have limited the accuracy and predictive power of parametric empirical models (Franklin, 1995; Guisan and Zimmermann, 2000). Artificial neural networks (e.g., BPN) offer the benefits of empirical modeling without adherence to parametric assumptions (Foody, 1995; Foody and Arora, 1997); potentially allowing for an improved empirical model when compared with methods rooted in traditional statistical theory (Pijanowski et al., 2002).

Studies comparing neural network and conventional (i.e., parametric) SDMs using the same dataset have been limited (Guisan and Zimmermann, 2000). Segurado and Araujo (2004) modeled amphibians and reptiles in Portugal using several techniques (e.g., neural networks, generalized linear models, generalized additive models) and found neural networks to consistently produce more accurate models, particularly when modeling high tolerance (i.e., low marginality) species like the gypsy moth. Mastrorillo et al. (1997) compared discriminant analysis and BPN to model several fish species' distribution and found BPN to produce 20% improved prediction accuracies when variables exhibited non-linear relationships. Manel et al. (1999) compared discriminant analysis, logistic regression, and a BPN to model the distribution of a Himalayan river bird and found the BPN to produce more accurate (i.e., overall map accuracy) predictions than logistic regression or discriminant analysis, but logistic regression outperformed BPN when validated using the AUC. Olden and Jackson (2001) compared BPN to logistic regression to model nine fish species using simulated Gaussian and linear response functions and BPN outperformed logistic regression by an average of 5.65% (overall map accuracy). They found BPN to have broad applicability to the study of ecological relationships for both exploratory and predictive purposes, particularly when species response curves are non-linear (Olden and Jackson, 2001).

2.3.1. *Multi-criteria evaluation*

Fuzzy multi-criteria evaluation (MCE) functions through a simple weighted linear combination of variables, where independent variable values indicate *probability* of occurrence of the modeled phenomenon to produce a map representing suitability of presence (Eastman et al., 1995). MCE allows for complete user control in that variable weights and relationships are predefined by the analyst. This facilitates the incorporation of ecological theory by forcing the explicit definition of variable weights and relationships, and allows modeling in the absence of presence/absence representations of species occurrences. However, because MCE requires *a priori* definition of variable weights and relationships, a thorough understanding of environment-species relationships is necessary for the sound prediction of distribution (Austin, 2002).

Methods have been developed to aid the identification of optimal variable weights, the most popular of which is the analytical hierarchy process (AHP) developed by Saaty (1980, 1987). The weights generated by the AHP are produced by means of the principal eigenvector of a pairwise matrix comparing the relative importance of input variables (Saaty, 1987; Eastman et al., 1995). The most common application of MCE techniques has been land allocation optimization for regional planning purposes (e.g., Lin et al., 1997; Corcoran et al., 1997; Antonie et al., 1997) though there are several examples of risk modeling implementations (e.g., Tkach and Simonvic, 1997; Duijm and Markert, 2002; Fuller et al., 2002). While MCE has not been applied to SDMs (i.e., realized distribution) it has been frequently applied to habitat suitability models (i.e., potential distribution) (e.g., Store and Jokimaki, 2003). It can, however, be argued that informal MCE is conducted each time a manager attempts to optimize resource distribution efficiency: factors known to contribute to gypsy moth presence are considered based on the experience of the manager and areas constituting the highest risk are identified. MCE is, in its simplest terms, a formalization and subsequent optimization of the process managers have typically employed.

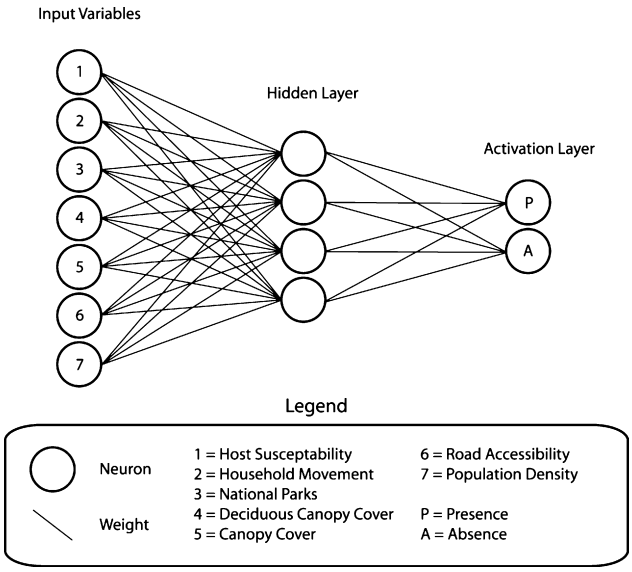

Fig. 1 – Conceptual model of the multi-layer perceptron neural network used in these analyses.

2.3.2. *Logistic regression*

Logistic regression is an empirical modeling technique used for prediction of a binary response variable (e.g., species presence/absence). Several parameter optimization techniques are available, the most popular of which is the maximum likelihood estimation procedure (Clark and Hosking, 1986; Eastman, 2006). Logistic regression assumes that outcomes are mutually exclusive and exhaustive, the dependent-predictor variable relationship is logistic, samples are random, and residual errors are independent (Eastman, 2006). Despite the frequent violation of these underlying assumptions, logistic regression has been the predominant method for probabilistic modeling of species distribution (Franklin, 1995) but has seen limited application for the prediction of gypsy moth distribution.

Liebhold et al. (1998) and Gribko et al. (1995) used logistic regression fit via maximum likelihood estimation to predict forest stand-level gypsy moth defoliation in Massachusetts. Gribko et al. (1995) found logistic regression based on trap counts to produce more realistic models to predict defoliation then three-dimensional Kriging of known defoliation. Liebhold et al. (1998) note, however, that logistic regression models offer no improvement over simple egg mass threshold methods (i.e., number of egg masses equals severity of defoliation).

2.3.3. *Back propagation neural network*

Multi-layer perceptrons trained using a back-propagation procedure (BPNs) are a form of feed-forward artificial neural network calibrated using a back propagation algorithm (Rumelhart et al., 1986). Based on a recursive learning procedure, the algorithm uses a gradient decent search to minimize model calibration error (Kanellopoulos and Wilkinson, 1997). BPNs have three primary components, an input layer, an output layer, and one or more hidden layers; each composed of a user-defined number of neurons. Output neurons represent the classes specified by the calibration data. Input variables and hidden layer neurons are randomly weighted and assigned membership to an output neuron. This process is repeated and the weights resulting in the lowest testing error are retained. Repeated iteratively, weights reach an approximately optimal solution for the partition of input variables into the specified output classes (i.e., presence-absence). Fig. 1 provides a conceptual model of the BPN used for the analyses presented here.

Unlike logistic regression, BPNs operate without parametric assumptions. Subsequently, they allow the characterization of models containing non-linear relationships and inherent dependence within, among, and between predictor variables without the explicit definition of those relationships (Lek and Guegan, 1999). This advantage can allow improved prediction accuracy compared to parametric techniques such as logistic regression (Manel et al., 1999). Neural networks like BPN represent a powerful, yet under explored, tool for integration into SDMs (but see Colasanti, 1991; Edwards and Morse, 1995; Fitzgerald and Lees, 1992, 1994; Lek and Guegan, 1999 and Guisan and Zimmermann, 2000).

BPNs are the most widely-used and, subsequently, the most extensively explored type of neural network algorithm in GIScience (Foody, 1995; Foody and Arora, 1997) but have seen limited application to species distribution modeling (Franklin, 1995; Guisan and Zimmermann, 2000). Typically used for the classification of remotely sensed data (e.g., Foody and Arora, 1997; Gopal and Woodcock, 1996; Foody, 1995), BPNs have been applied to SDMs (e.g., Fitzgerald and Lees, 1992, 1994; Lek et al., 1996; Mastrorillo et al., 1997; Lek and Guegan, 1999; Manel et al., 1999; Tourenq et al., 1999). BPNs have seen limited application to model the risk of invasive species (but see Vander Zanden et al., 2004). While several types of artificial neural networks likely have potential for application in empirical modeling, BPNs have been the primary type of algorithm implemented, likely due to software availability.

Table 2 – Descriptions of all modeled variables used in this study

Variable	Prediction association	Source	Data type	Data range	Model inclusion
Airport density	Introduction probability	U.S. Bureau of Transportation Statistics	Point	–	–
Distance from quarantined counties	Introduction probability	USDA APHIS	Continuous	0–2,567,769	MCE
Household movement from defoliated counties	Introduction probability	U.S. Census Bureau	Continuous	0–680.00	–
Household movement—quarantined counties	Introduction probability	U.S. Census Bureau	Continuous	0–242,652	MCE, LR, BPN
Infestation history	Introduction probability	USDA APHIS National Agricultural Pest Information System	Categorical	0–10	MCE
National and state parks	Introduction probability	USGS	Polygon	–	LR, BPN
Percent of population emigrated from defoliated counties	Introduction probability	U.S. Census Bureau/LandScan	Continuous	0–0.45	LR
Population density	Introduction probability	U.S. Department of Energy (LGPD, 2000)	Continuous	0–6,528	MCE, LR, BPN
Rail density	Introduction probability	U.S. Bureau of Transportation Statistics	Line	–	–
Road accessibility	Introduction probability	USGS	Line	–	MCE, LR, BPN
Host susceptibility	Establishment potential	USDA (Morin et al., 2005)	Continuous	0–37.59	LR, BPN
Percent canopy cover	Establishment potential	TREES (DeFries et al., 2000)	Continuous	0–75.74	LR, BPN
Percent coniferous canopy cover	Establishment potential	TREES (DeFries et al., 2000)	Continuous	0–72.93	MCE
Percent deciduous canopy cover	Establishment potential	TREES (DeFries et al., 2000)	Continuous	0–76. 97	MCE, LR, BPN
Minimum January temperature	Establishment potential	PRISM (Spatial Climate Analysis Service, 2004)	Continuous	−25.66–12.81	–

3. Methods

3.1. *Data*

Initial data selection was based on a decision support framework currently used by the USDA to aid the distribution of pheromone sampling traps (USDA, 2001). Variables were selected in an attempt to predict the two requirements for gypsy moth population establishment: introduction probability and/or establishment potential given an introduction. Table 2 presents a summary of the variables assessed for inclusion in all models examined in this study. To provide information on the accessibility of suitable host material to (anthropogenic) introduction, the variables road accessibility, airport density, and rail density were included. Because parks constitute a distinctive risk due to high numbers of visitors carrying potential gypsy moth egg mass substrate (e.g., campers, firewood, boats), the variables national and state parks were included. To account for introduction due to household migrations, the variables household movement from quarantined counties, percentage of population emigrated from quarantined counties, and movement from (gypsy moth) defoliated counties were included. To account for introductions related to daily activities (e.g., shipping of firewood, building materials), the variable distance from quarantined counties was included. To provide information on gypsy moth host availability and subsequent establishment potential, the variable host susceptibility was included. To account for species not included in the calculation of host susceptibility, which describes only "high preference" host species (Liebhold et al., 1997b), the variables percent tree canopy cover, percent deciduous tree canopy cover, percent coniferous canopy cover were included. Lastly, to account for gypsy moth diapause temperature requirements (see Allen et al., 1993 for a full description), minimum January temperature was included.

Calibration/validation data were provided through a 14 year record of gypsy moth pheromone trap counts per county for non-infested portions of the United States. Empirical models (logistic regression and BPN) were calibrated using data from 1991 to 2000 and validated using data from 2001 to 2004. Calibration years were selected to correspond to 2000 census migration data. The remainder of available reference data was used for validation. Four years (2001–2004) is longer than the gypsy moth establishment guidelines offered by the USDA (USDA, 2001).

For the calibration dataset, "Presence" was defined as five or more moths captured, in a given county, in 4 or more years between 1991 and 2000 (USDA APHIS). "Absence" was defined as counties in which traps were distributed but gypsy moth presence has never been recorded between 1991 and 2000. A total of 1695 training samples, 1486 absence and 209 presence, resulted from reference data filtering based on these criteria. For the validation dataset, "Presence" was defined as five or more moths captured in any year (2001–2004) and "Absence" was defined as zero moths captured in trapped counties. Validation data filtering resulted in 962 samples; 884 absence and 78 presence. Calibration and validation data filtering criteria

Table 3 – Description of weight structure resulting form analytical hierarchy process

Variable	Migration from infested	Distance from quarantine	Percent coniferous canopy cover	Percent deciduous canopy cover	Infestation history 1990–2000	Road access-ability	Population density
Migration from infested	1						
Distance from quarantine	1	1					
Percentage coniferous canopy cover	3	9	1				
Percentage deciduous canopy cover	4	9	2	1			
Infestation history 1990–2000	1/2	3	1/3	1/3	1		
Road access-ibility	1/3	3	1/8	1/8	1/2	1	
Population density	3	3	1/3	1/5	2	1	1
Resulting weight	0.0865	0.0358	0.2666	0.3656	0.0791	0.0550	0.1114

were reached through a great deal of consultation with USDA APHIS entomologists, in order to reflect the model's ultimate purpose, risk allocation.

Host susceptibility, a data layer estimating the total basal area/ha of "high" preference (Liebhold et al., 1997b) gypsy moth host species, was created by Morin et al. (2005) by kriging United States Forest Service (USFS) forest inventory analysis (FIA) plot data and limiting their presence to forest classes as indicated by the United States Geological Survey's National Land Cover Dataset (DeFries et al., 2000). All accessibility (i.e., road, airport and railroad) and parks (i.e., national and state parks) variables were calculated by placing a 1 km buffer around the target feature (to provide otherwise one-dimensional features area) and dividing the total buffer area within the county by the total area of the county. All other variables were aggregated through the averaging of all values within a given county.

3.2. *Models*

We compare three distinct types of distribution models in terms of their ability forecast gypsy moth establishment risk based on the sweet of environmental and anthropogenic variables described above: expert system (i.e., MCE), parametric (i.e., logistic regression), and non-parametric (i.e., BPN). Each of these models has an appropriate variable selection method associated with it: expert knowledge, statistical significance, and iterative selection based on training accuracy, respectively. Similarly, while logistic regression and BPN are empirical models and require training data, MCE requires no training data. Subsequently, in order to assess each model under best practice conditions, variable selection and model calibration were conducted on an individual basis.

3.2.1. *Multi-criteria evaluation*

The MCE model was constructed using a weighted linear combination calibrated using the AHP (Saaty, 1980; Saaty, 1987). The variables selected through literature review and expert consultation include: household movement from quarantined counties, percent deciduous canopy cover, percent coniferous canopy cover, road accessibility, population density, infestation history from 1990 to 2000, and distance from infested areas. AHP requires the creation of a square reciprocal matrix defining the relative importance of each variable to each other using a 9-point rating scale ranging from 1 (equal importance) to 9 (strongly more important) with ratings of less importance being expressed as the reciprocal (i.e., strongly less important would be expressed as 1/9). By definition the diagonal entries are all equal to 1 (variables are equally important when compared to themselves) and the rating in any position *i,j* will be the reciprocal of that in position *j,i*. The Principal Eigenvector of this matrix then yields the importance weights of the variables (Table 3).

An important source of feedback in the AHP is the evaluation of the Consistency Ratio which expresses the degree to which the ratings form a consistent set of relationships. Saaty (1977) has shown that for a perfectly consistent set of ratings, the Principal Eigenvalue will be equal to the order of the matrix. This leads to a simple measure of departure from

Table 4 – Description of the factors included in the multi-criteria evaluation model

Factors	Function	A	B	C	D	Weight
Migration from infested	Sigmoid increasing	0	100			0.0865
Distance from quarantine	Sigmoid increasing	0	100			0.0358
Percentage coniferous canopy cover	Sigmoid symmetrical	54	449	449	56000	0.2666
Percentage deciduous canopy cover	Sigmoid increasing	0	Max			0.3656
Infestation history 1990–2000	Sigmoid Increasing	0	14			0.0791
Road accessibility	Direct					0.0550
Population density	Sigmoid decreasing			0	2000	0.1114

this ideal condition known as the Consistency index:

$$\mathrm{CI} = \frac{\lambda_{\max} - n}{n - 1} \quad (1)$$

where $\lambda_{\max}$ is the Principle Eigenvalue of the reciprocal matrix and n is the order of the ratings matrix. The Consistency Ratio (CR) is then the ratio of that index to the average CI for a large set of randomly generated ratings. Saaty (1977) suggested that when the CR exceeds 0.1 the ratings are inconsistent and should be re-generated. The AHP Consistency Ratio in this study was 0.08 indicating that the weights of variables were determined from an acceptably consistent set of ratings.

Table 4 provides a complete description of variable preparation. Variables were prepared and weighted via AHP based on information from the Gypsy Moth Manual (USDA, 2001), supporting literature (e.g., Leonard, 1981; Carter et al., 1994; Shaub et al., 1995; Sharov, 1996; Nealis et al., 2001) and advice from USDA APHIS and USDA Forest Service research entomologists.

3.2.2. *Logistic regression*

Using logistic regression, presence-absence reference counties served as the dependent variable and indicator variables provided independent variables. Through a stepwise method, the following variables were selected for inclusion in the model at $p \leq 0.05$: host susceptibility, household movement from quarantined counties, national parks, percentage deciduous canopy cover, percentage canopy cover, road accessibility, population density, and percentage of population emigrated from quarantined counties. All available training data (i.e., 209 presence and 1486 absence) were used for calibration.

3.2.3. *Back-propagation neural network*

Through an iterative selection process, the following variables were selected: host susceptibility, household movement from quarantined counties, national parks, percentage deciduous canopy cover, percentage canopy cover, road accessibility, and population density (Fig. 1). Since traditional automated stepwise methods based on significance testing are based on an assumption of normality and could subsequently reject explanatory (non-normal) variables as insignificant, a subjective iterative selection method was adopted. All possible variable combinations were modeled and the suite of variables producing the lowest root mean square testing error (i.e., training accuracy) was retained.

BPN requires the specification of several parameters, arguably the most influential of which is the learning rate (Kavzoglu and Mather, 2003). This parameter determines the maximum weight adjustment at each iteration. The implementation of BPN used for these analyses allows for the automatic adjustment of the learning rate based on root mean square error fluctuations over several iterations. The automatic calculation of learning rate allows for a reasonable approximation of an optimal setting while reducing the amount of trial and error necessary for parameter selection and the likelihood of overtraining. A single hidden layer with four nodes was used. A momentum factor of 0.5 and sigmoid constant of 1.0 were found to be optimal through a trial and error procedure. All available training information was used to create equally proportioned training and testing samples (Tourenq et al., 1999); 208 samples for training and 208 samples for testing; 104 for each presence and absence. To allow the network to 'self-check' at each iteration, training data presented to the network are sampled into 50% training, and 50% testing, with presence an absence in equal proportion. These parameters allowed for significant convergence after 10,000 iterations.

3.3. *Model assessment*

Spatially explicit models of gypsy moth infestation risk were validated using presence/absence information from 2001 to 2004, where presence was defined as five or more moths captured in any given year, based on two criteria: AUC and a simple threshold assessment. AUC provides a threshold and prevalence independent measure of a model's predictive power which permits model validation independent of distortions and potential bias introduced by dichotomization (Fielding and Bell, 1997). AUC requires the rank ordering of a suitability image and thresholding of that rank ordered image at a user specified number of intervals to produce a boolean map that is then compared to the boolean map of true presence (Eastman, 2006; Pontius and Schneider, 2001). True positives and false positives are plotted; the area between the plotted line and random (i.e., equal true and false positives), as a proportion of the total area above random, is the area under the curve AUC statistic. The (AUC) was calculated as:

$$\mathrm{AUC} = \sum_{i=1}^{n} [x_{i+1} - X_i] x \left[\frac{y_i + (y_{i+1} - y_i)}{2} \right] \quad (2)$$

where x_i is the rate of false positives for threshold i, y^i is the rate of true positives for threshold i, and $n+1$ is the number of thresholds. One hundred and one thresholds were used (i.e., $n = 100$) for these analyses.

Simple threshold assessment refers to a similar process excepting the rank ordering of presence suitability values prior to thresholding; eliminating potential artifacts associ-

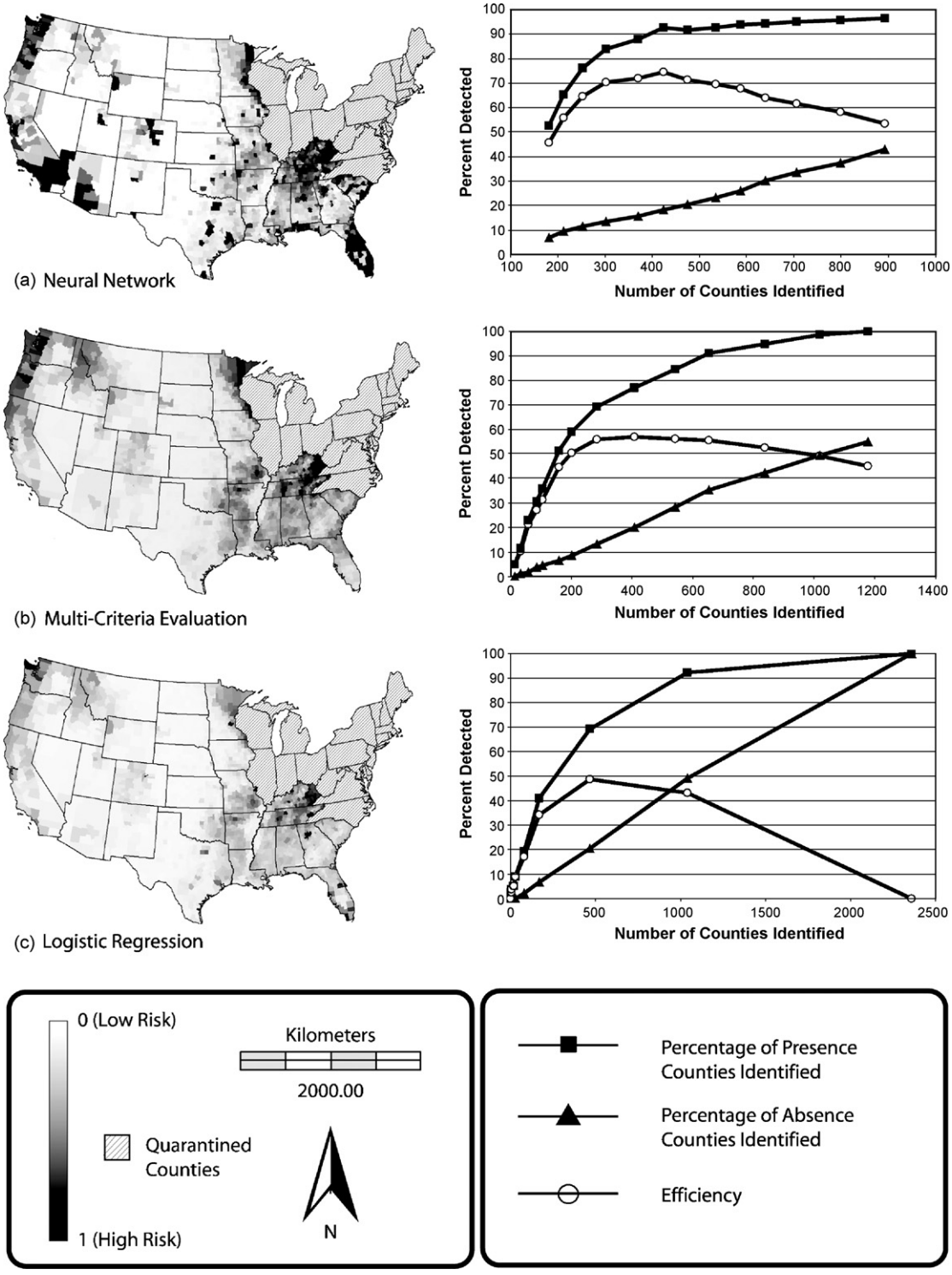

Fig. 2 – Model projections and efficiency for (a) neural network, (b) multi-criteria evaluation, and (c) logistic regression. Efficiency graphs compare percentage of known infestation detected by the model to the number of counties identified as infested by the model.

ated with the rank order process. Rank ordering can lead to pixels of the same suitability (i.e., risk) value being calculated in different thresholds; potentially introducing bias into both ROC curves and AUC. Raw value (0–1) models were thresholded at 0.05 intervals to determine true and false positives and to ultimately allow the identification of an approximate threshold of maximum efficiency for each model. Maximum efficiency is calculated as:

$$\text{Maximum efficiency} = \max_{i=20} [x_i - y_i] \quad (3)$$

where x_i is the percentage of true positives at threshold i, and y_i is the percentage of false positives at threshold i.

4. Results and discussion

Of the three models assessed, BPN provided the most accurate predictions based on independent forecast validation using 2001–2004 trap counts producing a AUC of 0.93, followed by MCE (AUC = 0.92) and logistic regression (AUC = 0.86) (Fig. 2). BPN, with optimized parameters and a training accuracy of 94.2%, produced a model with a maximum efficiency of 74.5%. Maximum efficiency was realized at a threshold of 0.65, correctly identifying 92.7% of infestations (424 counties or 769,034 km^2 total) from 2001 to 2004 while falsely identifying 18.2% of areas known to be non-infested. Fig. 2a describes omission errors, commission errors, and efficiency for the BPN model in terms of the number of counties identified as a trapping priority.

MCE produced a model (Fig. 2b) with a maximum efficiency of 64.6%. Maximum efficiency was realized at a threshold of 0.55, correctly identifying 69.2% of infestations (408 counties or 556,230 km^2 total) from 2001 to 2004 while falsely identifying 13.2% of areas known to be non-infested. Fig. 2b describes omission errors, commission errors, and efficiency for the MCE model in terms of the number of counties identified for trapping priority.

Logistic regression produced the following model (Fig. 2c):

$$\begin{aligned}\text{Logit(presence/absence)} = {} & -4.56 - 0.12A + 0.00B - 1.75C \\ & +0.08D + 0.03E + 3.29F + 0.01G \\ & -0.25H \end{aligned} \quad (4)$$

where A is host susceptibility, B is household movement from quarantined counties, C is national parks, D is percentage deciduous canopy cover, E is percentage canopy cover, F is road accessibility, G is population density, and H is the percentage of population emigrated from quarantined counties. Maximum efficiency (48.73%) was realized at a threshold of 0.4, correctly identifying 69.2% of infestations (468 counties or 770,997 km^2 total) from 2001 to 2004 while identifying 20.47% of areas known to be non-infested. Fig. 2c describes omission errors, commission errors, and efficiency for the logistic regression model in terms of the number of counties identified as a trapping priority.

MCE, logistic regression, BPN, and persistence (i.e., if the prediction were persistence at year 2000 locations) ROC curves are summarized in Fig. 3. Note that the BPN curve rises (i.e., increase in true positives) quickly, in comparison to the MCE curve, but levels off, while the MCE curve rises to 100%. This indicates that, in comparison to BPN, MCE systematically overpredicts presence below the 90% true positive threshold. MCE is therefore the modeling method of choice when unlimited trapping resources are available, but BPN produces the most accurate sampling scheme to trap up to 90% of potential infestations.

The inability of the BPN model to accurately characterize gypsy moth distribution beyond the 90th percentile suggests that training data do not exhaustively characterize the range

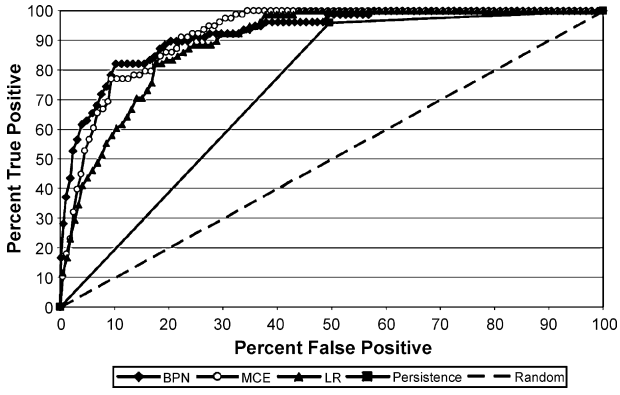

Fig. 3 – Relative operating characteristics for back propagation network, multi-criteria evaluation, and logistic regression compared to relative operating characteristic for persistence.

of potential host sites; suggesting that there are potential host ranges that are not being sampled. The ability of MCE to characterize risk beyond the 90th percentile suggests that the variables included contain the information necessary for the prediction of infestation beyond the 90th percentile and subsequently reinforce the suggestion that the training data may not fully capture the variance of the gypsy moth potential infestation range. The gypsy moth's demonstrated ability to adapt to less than ideal environments (USDA, 2001) may explain the training data limitation; trap distribution is based on current understanding of potential gypsy moth range, however, gypsy moth adaptation capabilities are not fully documented or understood and there is potential for gypsy moth to adapt to areas outside its commonly accepted range (Allen et al., 1993). In empirical models (e.g., BPN) it is assumed that sample data characterize the variance of the population, which APHIS trap count data may not.

Despite being calculated in very different ways, the three models exhibit large areas of location agreement (e.g., Pacific Northwest and Appalachia). Areas consistently identified as high risk are predominantly areas exhibiting high quantities of host material. MCE weights (Table 3) reveal that deciduous and coniferous canopy covers play a significant role in the calculation of the MCE model. Logistic regression exhibits a similar pattern of apparent dependence on canopy cover (i.e., host) variables. BPN, however, follows a pattern combining canopy cover and household movement. Many areas of disagreement between the models are locations which, during the time period of the study, experienced high volumes of immigration, an important consideration when accounting for introduction probability. The inability of logistic regression to characterize household movement data indicates that BPN provides an improved method for the integration of non-normal anthropogenic variables when compared to logistic regression.

The three models identify highly variable quantities of high-risk areas. BPN predicts the highest quantity of high-risk areas, followed by MCE and logistic regression respectively. The MCE model risk quantity and distribution is a product of expert derived response curves while BPN and logistic

regression are derived empirically from the same independent variables. Both empirical models (i.e., BPN, logistic regression) are influenced more profoundly by household movement than the MCE model; indicating that household movement is more important to prediction of gypsy moth distribution than previously thought by experts and ultimately resulting in lower risk values in many high migration areas by the MCE model (e.g., Fulton County, GA, and Denver County, CO) when compared to the empirical models. There are, however, significant differences between the two empirical models. The BPN model identifies a greater quantity of high-risk areas than the logistic regression model. This can be explained by the function used by each of the models: parametric assumptions inherent to the logistic regression model do not allow it to accurately characterize the non-linear relationships between several of the anthropogenic variables and gypsy moth establishment risk.

Three anthropogenic variables (i.e., migration from infested locations, road accessibility, and population density) were used in the calculation of all three models. In all models anthropogenic variables play a significant role in the calculation of gypsy moth distribution; indicating that the inclusion of anthropogenic variables in the calculation of SDMs can contribute significantly to the accurate and robust prediction of species distribution.

5. Conclusions

This research compared the accuracy of expert, parametric and neural network modeling techniques for the integration of anthropogenic and ecological variables in support of invasive species risk forecasting. The BPN and MCE algorithms provided comparably accurate predictions (AUC = 0.93 and 0.92, respectively) of gypsy moth infestation, both significantly more accurate than logistic regression (AUC = 0.86). Unlike MCE however, BPN produced a prediction independent of *expert* knowledge. This finding demonstrates that BPN can elucidate factors contributing to the introduction (i.e., predictor–response variable relationships) of invasive species for which variable relationships are not fully understood. The integration of anthropogenic variables enabled the production of an accurate risk-model providing insight into the impact of anthropogenic activities (e.g., household moves) on the risk of gypsy moth infestation in the US. For the prediction of gypsy moth infestation risk, household movement data provided the single most powerful predictor (variable) of gypsy moth presence. Further, BPN provided a robust technique for integrating variables representing anthropogenic interaction and ecological properties that are capable of accurately predicting pest-risk without *a priori* understanding of predictor/response variable relationships. This method can be applied to develop risk models to inform managers of factors contributing to the establishment of invasive species (fauna and flora) in North America and other environments. The models developed through this research directly inform mitigation strategies of APHIS managers.

The integration of anthropogenic variables into species distribution modeling remains an open avenue for research, particularly with regard to predictive vegetation modeling. Many flora are altered or disturbed by anthropogenic activities; the inclusion of anthropogenic data therefore, has potential to improve predictions and further understanding of the relationship between the modeled species and anthropogenic activities.

Acknowledgements

This research was funded through a cooperative agreement between USDA APHIS and Clark Labs, Clark University #05-8100-0988-CA. Research was also made possible by support and technical expertise from Clark Labs; an organization dedicated to the research and development of geospatial technologies for effective and responsible decision making for environmental management, sustainable resource development and equitable resource allocation. All processes conducted in the course of this research were complete in the GIS and image processing software Idrisi Andes 15.0.

REFERENCES

Allen, J.C., Foltz, J.L., Dixon, W.N., Liebhold, A.M., Colbert, J.J., Regniere, J., Gray, D.R., Wilder, J.W., Christie, I., 1993. Will the gypsy moth become a pest in Florida? Florida Entomologist 76 (1), 102–113.

Antonie, J., Fischer, G., Makowski, M., 1997. Multiple criteria land use analysis. Appl. Mathematics Comput. 85, 195–215.

Araujo, M.B., Williams, P.H., 2000. Selecting areas for species persistence using occurrence data. Biol. Conservation 96, 331–345.

Araujo, M.B., Williams, P.H., Fuller, R.J., 2002. Dynamics of extinction and the selection of nature reserves. Proc. R. Soc. Lond. B269, 1971–1980.

Austin, M.P., 1980. Searching for a model for use in vegetation analysis. Vegetation 42, 11–21.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecol. Modell. 157, 101–118.

Austin, G., Thomas, C., Houston, D., Thompson, D.B.A., 1996. Predicting the spatial distribution of buzzard *Buteo buteo* nesting areas using a Geographical Information System and remote sensing. J. Appl. Ecol. 33, 1541–1550.

Bergman, B.L., Chandler, M.D., Locklear, A., 2000. The economic impact of invasive species to wildlife services' cooperators. In: Proceedings of the Third NWRC Special Symposium, August 1–3, 2000, Fort Collins, CO.

Byers, J.E., Reichard, S., Randall, J.M., Parker, I.M., Smith, C.S., Lonsdale, W.M., Atkinson, I.A.E., Seastedt, T.R., Williamson, M., Chornesky, E., Haynes, D., 2002. Directing research to reduce the impacts of nonindigenous species. Conservation Biol. 16, 630–640.

Carter, J.L., Ravlin, F.W., Gray, D.R., Carter, M.R., Coakley, C.W., 1994. Foliage presence and absence effect on gypsy moth (Lepidoptera: Lymantriidae) egg mass sample counts and the probability of exceeding action thresholds with foliage present. J. Econ. Entomol. 87 (4), 1004–1007.

Clark, W.A.V., Hosking, P.L., 1986. Statistical Methods for Geographers. Wiley, New York, NY.

Colasanti, R.L., 1991. Discussions of the possible use of neural network algorithms in ecological modeling. Binary 3, 13–15.

Corcoran, K., Dent, B., Smith, J., Lara, P., 1997. Location of optimal areas of development of an alternative livestock species: the cashmere goat. In: Laker, J., Milne, J. (Eds.), Proc. Livestock

Systems in Rural Development Network Conferences. Athens, pp. 67–72.
Cumming, G.S., 2002. Comparing climate and vegetation as limiting factors for species ranges of African ticks. Ecology 83, 255–268.
DeFries, R., Hansen, M., Townshend, J.R.G., Janetos, A.C., Loveland, T.R., 2000. Continuous Fields 1 KM Tree Cover. The Global Land Cover Facility, College Park, Maryland.
Duijm, N.J., Markert, F., 2002. Assessment of technologies for disposing explosive water. J. Hazard. Mater. A90, 137–153.
Eastman, J.R., 2006. Idrisi 15.0 User's Guide. Clark Labs, Worcester, MA.
Eastman, J.R., Jin, W., Kyem, P.A.K., Toledano, J., 1995. Raster procedures for multi-criteria/multiobjective decisions. Photogrammetry Remote Sensing 61 (5), 539–547.
Eastman, J.R., Kyem, P.A.K., Toledano, J., Jin, W., 1993. Explorations in Geographic Systems Technology, vol. 4. GIS and Decision Making, Geneva, Switzerland, UNITAR.
Edwards, M., Morse, D.R., 1995. The potential for computer aided identification in biodiversity research. Trends Ecol. Evol. 10, 153–158.
Ferrier, S., 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? Syst. Biol. 51 (2), 331–363.
Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conservation 24 (1), 38–49.
Fitzgerald, R.W., Lees, B.G., In: American Society of Photogrametry and Remote Sensing (Eds), 1992. The application of neural networks to the floristic classification of remote sensing and GIS data in complex terrain. In: Proceedings of the XVII Congress ASPRS, Bethesda, MD, pp. 570–573.
Fitzgerald, R.W., Lees, B.G., 1994. Assessing the classification accuracy of multisource remote sensing data. Remote Sensing Environ. 47, 362–368.
Foody, G.M., 1995. Land-cover classification by an artificial neural network with ancillary information. Int. J. Geographical Inf. Syst. 9 (5), 527–542.
Foody, G.M., Arora, M.K., 1997. Evaluation of some factors affecting the accuracy of classification by an artificial neural network. Int. J. Remote Sensing 18, 799–810.
Franklin, J., 1995. Predictive vegetation mapping: geographic modeling of biospatial pattern in relation to environmental gradients. Prog. Phys. Geography 19 (4), 474–499.
Fuller, D., Jeffe, M., Williamson, R.A., James, D., 2002. Satellite remote sensing and transportation lifelines: safety and risk analysis along rural southwest roads. In: Pecora 15/Land Satellite Information IV/ISPRS Commission 1/FIEOS 2002 Proceedings.
Gahegan, M., 2003. Is inductive machine learning just another wild goose (or might it lay the golden egg)? Int. J. Geographic Inf. Sci. 17 (l), 69–92.
Gerardi, M.H., Grimm, J.K., 1979. The History, Biology, Damage, and Control of the Gypsy Moth, *Porthetria dispar* (L.). Associated University Presses, Cranberry, New Jersey.
Gopal, S., Woodcock, C.E., 1996. Remote sensing of forest change using artificial neural networks. IEEE Trans. Geoscience Remote Sensing 34, 398–404.
Gottschalk, K.W., 1993. Silvicultural guidelines for forest stands threatened by the gypsy moth. U.S. Department of Agriculture Forest Service General Technical Report NE-171.
Gribko, L.S., Liebhold, A.M., Hohn, M.E., 1995. Model to predict gypsy moth (Lepidoptera: Lymantriidae) defoliation using kriging and logistic regression. Environ. Entomol. 24 (3), 529–537.
Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C.C., Aspinall, R., Hastie, T., 2006. Making better biogeographical predictions of species' distributions. J. Appl. Ecol. 43, 386–392.
Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009.
Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Modell. 135, 147–186.
Herrick, O.W., Gasner D.A., 1986. Rating forest stands for gypsy moth defoliation. Research Paper NE-583. Radnor, PA: USDA Forest Service.
Kanellopoulos, I., Wilkinson, G.G., 1997. Strategies and best practice for neural network image classification. Int. J. Remote Sensing 18 (4), 711–725.
Kavzoglu, T., Mather, P.M., 2003. The use of backpropogating artificial neural networks in land cover classification. Int. J. Remote Sensing 24 (23), 4907–4938.
2000. Landscan Global Population Database. Oak Ridge National Laboratory, Oak Ridge, TN, Available at http://www.ornl.gov/landscan.
Lek, S., Guegan, J.F., 1999. Artificial neural networks as a tool in ecological modeling, an introduction. Ecol. Modell. 120, 65–73.
Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modeling non linear relationships in ecology. Ecol. Modell. 90, 39–52.
Leonard, D.E., 1981. Bioecology of the Gypsy Moth. In: Doane, C.C., McManus, M.L. (Eds.), The Gypsy Moth: Research Toward Integrated Pest Management. USDA Technical Bulletin 1584, Washington DC, pp. 9–29.
Leuschner, W.A., Young, J.A., Walden, S.A., Ravlin, F.W., 1996. Potential benefits of slowing the gypsy moth's spread. Southern J. Appl. Forestry 20, 65–73.
Liebhold, A., Mastro, V., Schaefer, P.W., 1989. Learning from the legacy of Leopold Trouvelot. Bull. Entomol. Soc. Am. 35, 20–21.
Liebhold, A.M., Gottschalk, K.W., Luzader, E.R., Mason, D.A., Bush, R., Twardus, D.B., 1997a. Gypsy Moth in the United States: An Atlas. USDA Forest Service General Technical Report NE-233.
Liebhold, A.M., Gottschalk, K.W., Mason, D.A., Bush, R.R., 1997b. Forest susceptibility to the gypsy moth. J. Forestry 95, 20–24.
Liebhold, A.M., Halverson, J., Elmes, G., 1992. Quantitative analysis of the invasion of gypsy moth in North America. J. Biogeography 19, 513–520.
Liebhold, A., Luzader, E., Reardon, R., Roberts, A., Ravlin, F.W., Sharov, A., Zhou, G., 1998. Forecasting gypsy moth (Lepidoptera: Lymantriidae) defoliation with a geographical information system. J. Econ. Entomol. 91 (2), 464–472.
Liebhold, A.M., Luzader, E., Reardon, R., Bullard, A., Roberts, A., Ravlin, W., Delost, S., Spears, B., 1996. Use of geographic information systems to evaluate regional treatment effects in a gypsy moth (Lepidoptera: Lymantriidae) management program. J. Econ. Entomol. 89 (5), 1192–1203.
Liebhold, A.M., Macdonald, W.L., Bergdahl, D., Mastro, V.C., 1995. Invasion by exotic pests: a threat to forest ecosystems. Forest Sci. Monogr. 30, 49.
Lin, H.Q., Wan, X., Li, J., Chen, Y., Kong, 1997. GIS-based multicriteria evaluation for investment environment. Environ. Plan. B: Plan. Des. 24, 403–414.
Manel, S., Dias, J., Ormerod, S.J., 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecol. Modell. 120, 337–347.
Mastrorillo, S., Lek, S., Dauba, F., Belaud, A., 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. Freshwater Biol. 2, 237–246.
Morin, R.S., Liebhold, A.M., Luzador, E.R., Lister, A.J., Gottschalk, K.W., Twardus, D.B., 2005. Mapping Host-Species Abundance of Three Major Exotic Forest Pests. Research Paper NE-726. Newtown Square, PA: USDA Forest Service.
Nealis, V., Regniere, J., Gray, D. Modeling Seasonal Development of the Gypsy Moth in a Novel Environment for Decision Support of an Eradication Program. Pp 124–132. In: Liebhold,

A.M., McManus, M.L., Otvos, I.S., Fosbroke, S.L.C. (Eds.) 2001. Proceedings: integrated pest management and dynamics of forest defoliating insects. August 15–19, 1999, Victoria B.C. General Technical Report NE-277. Newtown Square, PA: USDA Forest Service.

Olden, J.D., Jackson, D.A., 2001. Fish-habitat relationships in lakes: gaining predictive and explanatory insight using artificial neural networks. Trans. Am. Fisheries Soc. 130, 878–897.

Osborne, P.E., Alonso, J.C., Bryant, R.G., 2001. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. J. Appl. Ecol. 38, 458–471.

Ozesmi, S.L., Tan, C.O., Ozesmi, U., 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Modell. 195, 83–93.

Pijanowski, B.C., Brown, D.G., Shellito, B.A., Manik, G.A., 2002. Using neural networks and GIS to forecast land use changes: a land transformation model. Comput. Environ. Urban Syst. 26 (6), 553–575.

Pimentel, D., Lach, L., Zuniga, R., Morrison, D., 1999. Environmental and Economic Costs Associated with Non-Indigenous Species in the United States. Cornell University, College of Agriculture and Life Sciences, Ithaca, New York, Available: http://www.news.cornell.edu/releases/Jan99/species_costs.html.

Pontius Jr., R.G., Schneider, L., 2001. Land-use change model validation by a ROC method for the Ipswich watershed, Massachusetts, USA. Agric. Ecosyst. Environ. 85 (1–3), 239–248.

Rumelhart, D., Hinton, G., Williams, R., 1986. Learning internal representations by error propogation. In: Rumelhart, D.E., McClelland, J.L. (Eds.), Parallel Distributed Processing: Explorations in the Microstructures of Cognition, vol. 1. MIT Press, Cambridge, pp. 318–362.

Saaty, T.L., 1977. A scaling method for priorities in hierarchical structures. J. Math. Psychol. 15 (3), 234–281.

Saaty, T.L., 1980. The Analytical Hierarchy Process. McGraw Hill, New York.

Saaty, R.W., 1987. The analytic hierarchy process—what it is and how it is used. Math. Model. 9 (3), 161–176.

Segurado, P., Araujo, M., 2004. An evaluation of methods for modeling species distributions. J. Biogeography 31, 1555–1568.

Shaub, L.P., Ravlin, F.W., Gray, D.R., Logan, J.A., 1995. Landscape framework to predict phelogical events for gypsy moth (Lepidoptera: Lymantriidae) management programs. Environ. Entomol. 24 (1), 10–18.

Sharov, A., 1996. Modeling insect dynamics. In: Korpilahti, E., Mukkela, H., Salonen, T. (Eds.), Caring for the forest: research in a changing world. Congress Report, vol. II., IUFRO XX World Congress, 6–12 August 1995, Tampere, Finland. Gummerus Printing, Jyvaskyla, Finland, 293–303.

Sharov, A.A., Liebhold, A.M., Roberts, E.A., 1997. Methods for monitoring the spread of gypsy moth (Lepidoptera: Lymantriidae) populations in the Appalachian Mountains. J. Econ. Entomol. 90, 1259–1266.

Spatial Climate Analysis Service, 2004. PRISM Group. Oregon State University, http://www.prismclimate.org.

Stohlgren, T.J., Schnase, J.L., 2006. Risk analysis for biological hazards: what we need to know about invasive species. Risk Anal. 26 (1), 163–173.

Store, R., Jokimaki, J., 2003. A GIS-based multi-scale approach to habitat suitability modeling. Ecol. Modell. 169 (1), 1–15.

Suárez-Seoane, S., Osborne, P.E., Alonso, J.C., 2002. Large-scale habitat selection by agricultural steppe birds in Spain: identifying species-habitat responses using generalized additive models. J. Appl. Ecol. 39, 755–771.

Tkach, R.J., Simonvic, S.P., 1997. A new approach to multi-criteria decision making in water resources. J. Geographic Inf. Decis. Anal. 1 (1), 25–43.

Tourenq, C., Aulagnier, S., Mesleard, F., Durieux, L., Johnson, A., Gonzolez, G., Lek, S., 1999. The use of artificial neural networks for predicting rice crop damage by greater flamingos in the Camargue, France. Ecol. Modell. 120, 349–358.

USDA Forest Service, 1992. Forest insect and disease conditions in the United States 1991. U.S. Department of Agriculture Forest Service Forest Pest Management, Washington DC, USA.

United States Department of Agriculture Animal and Plant Heath Inspection Service Plant, Protection and Quarantine. Gypsy Moth Program Manual. 2001.

Vander Zanden, M.J., Olden, J.D., Thorne, J.H., Mandrak, N.E., 2004. Predicting occurrences and impacts of smallmouth bass introductions in north temperate lakes. Ecol. Appl. 14, 132–148.